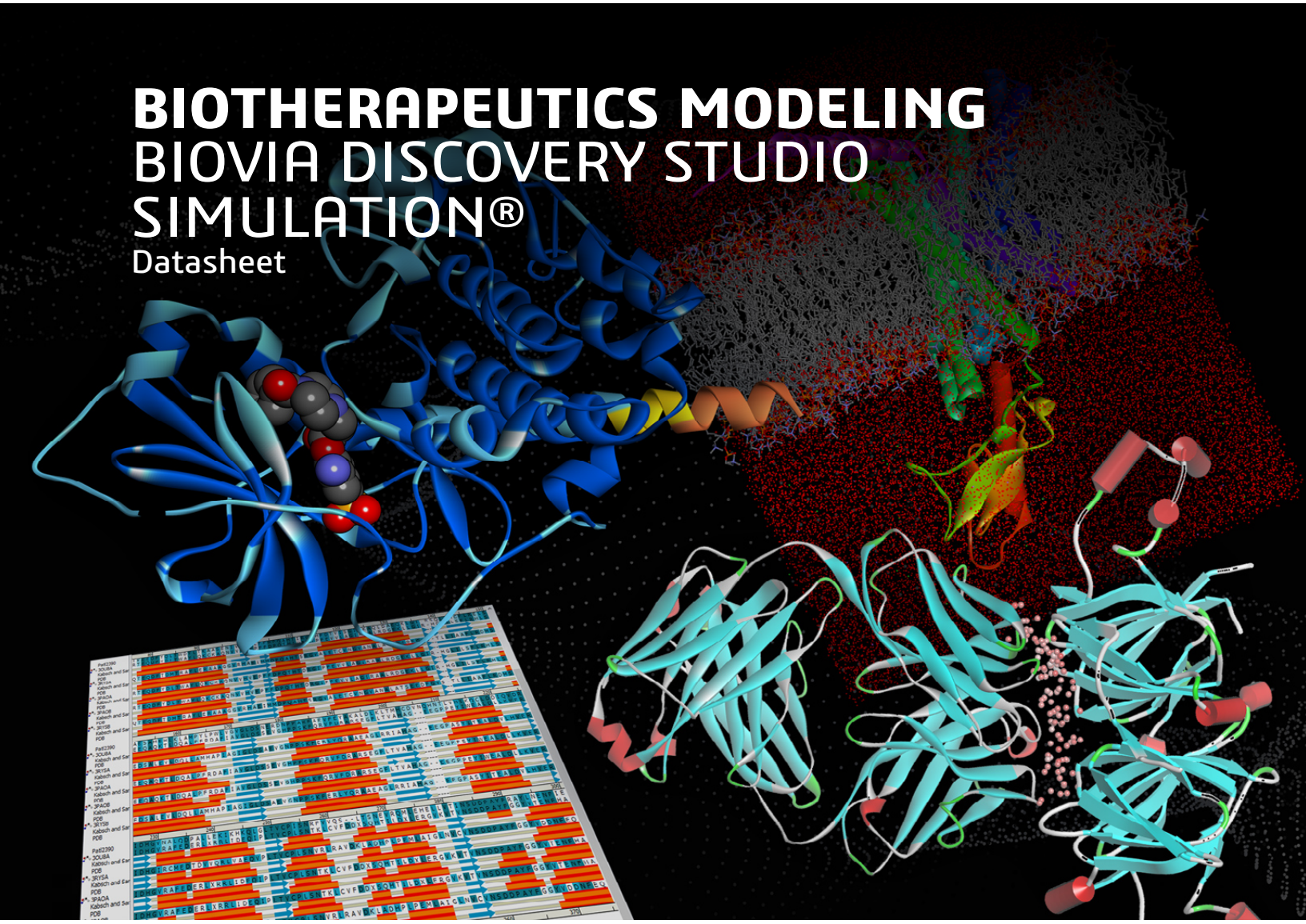
BIOVIA

3DEXPERIENCE®

BIOTHERAPEUTICS MODELING
BIOVIA DISCOVERY STUDIO SIMULATION®
Datasheet



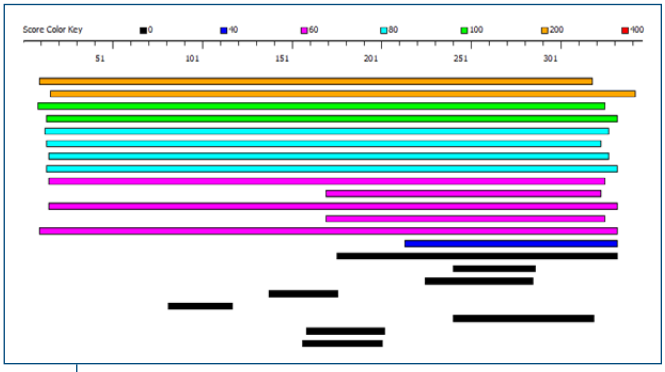
COMPREHENSIVE PROTEIN MODELING

Determining the three-dimensional structure and properties of macromolecules, such as enzymes, receptors, antibodies, DNA, or RNA, is a fundamental component of a wide range of research activities. For example, predicting the location and characteristics of binding sites or optimizing the stability and selectivity of therapeutic biologics all require access to precise, accurate molecular models. BIOVIA Discovery Studio Simulation delivers a comprehensive portfolio of AI models and market-leading, validated molecular modeling methods that assist with every aspect of macromolecule-based research. This datasheet outlines the capabilities of Discovery Studio Simulation for biotherapeutics modeling.

DASSAULT SYSTEMES

SEQUENCE ANALYSIS

- Perform sequence similarity searches using BLAST and PSI-BLAST against local or NCBI databases
- Perform a range of feature and motif predictions and biophysical property calculations on protein sequences
- Predict sites prone to post-translational modifications (PTM)
- Align sequences quickly and accurately using multiple sequence-alignment algorithms
- Determine relationships between sequences and structural conservation of amino acids with phylogenetic and Evolutionary Trace analysis tools



	Title/Description	Accession	Identity	Sequence	Alignment	Bit Score	E-value	Positive	Resolution	SCOP	Ligand	Organism
1	Adenosine deami...	3pao_A	44	313	310	266.929	9.04392e-88	63	2.493		A:ADE328,...	Pseudomonas aer...
2	Adenosine deami...	3rys_A	40	335	328	252.677	7.34169e-82	61	2.601	c.1.9.0	A:ADE345,...	Paenarthrobacter...
3	Adenosine deami...	4qxw_A	24	369	334	106.686	6.15386e-26	46	1.3		A:ZN401	Burkholderia ambi...
4	Adenosine deami...	1vfl_A	24	349	338	102.449	1.85466e-24	47	1.8	c.1.9.1	A:ZN501	Bos taurus
5	Adenosine deami...	6n91_B	24	334	326	99.3673	2.04024e-23	44	2.05	c.1.9.0	B:CXS403,...	Vibrio cholerae O...
6	Adenosine deami...	3mvl_A	25	349	331	98.5969	4.14232e-23	46	1.6	c.1.9.1	A:GOL902,...	Mus musculus
7	Adenosine deami...	6n9m_A	25	331	324	93.2041	2.5217e-21	44	1.449	c.1.9.0	A:CA404,A...	Salmonella enteri...
8	Adenosine deami...	3ior_A	23	360	339	91.6633	1.25973e-20	45	1.52	c.1.9.1	A:3D1501,...	Homo sapiens
9	adenosine deami...	2amx_A	25	364	328	78.9518	3.32681e-16	47	2.02	c.1.9.1	A:CO1000,...	Plasmodium yoelii
10	Adenosine deami...	3ewd_A	29	364	159	76.2554	2.46066e-15	54	1.9	c.1.9.1	A:MCF377,...	Plasmodium vivax
11	Adenosine deami...	6il7_A	23	355	335	73.1738	3.12261e-14	44	2.48	c.1.9.1	A:HPA402,...	Plasmodium falcip...
12	adenosine deami...	2pqf_A	29	359	165	72.4034	5.58789e-14	52	1.89	c.1.9.1	A:ADN501,...	Plasmodium vivax
13	Adenosine/AMP d...	6ijn_A	24	341	337	71.2478	1.03391e-13	43	1.66	c.1.9.0		Arabidopsis thalia...
14	Adenosine deami...	3lgd_B	29	482	126	41.9726	0.00051262	50	2		B:NAG750,...	Homo sapiens
15	AMP deaminase 2...	6hu6_A	22	614	171	40.0466	0.00255754	40	2.33		A:ZN901	Homo sapiens
16	Yeast Guanine De...	6oh9_A	40	452	47	32.7278	0.497529	59	1.75		A:ZN501	Saccharomyces c...
17	AMP deaminase |...	2a3l_A	30	616	60	32.3426	0.589059	46	3.34	c.1.9.1	A:CF5841,...	Arabidopsis thalia...
18	GUANINE PHOSP...	1dqn_A	32	230	46	31.187	0.919487	54	1.75	c.61.1.1	A:IMU300,...	Giardia intestinalis
19	Dihydrolipoyllysin...	6zzn_A	37	224	35	31.187	1.10893	48	1.5			Mycobacterium tu...
20	GENERAL ODORA...	2wcj_A	28	141	78	29.6462	2.13762	51	1.4	a..39.2.0	A:M21114...	Bombyx mori
21	O-acetyl-ADP-rib...	4j5r_A	39	141	43	28.4906	4.52546	58	1.25		A:AJR201	Homo sapiens
22	TRNA ENDONUCL...	1a79_A	31	171	45	28.8758	4.77243	62	2.28	c.52.2.1...	A:AU4	Methanocaldococ...

Figures 1 & 2: BLAST hits in a map and table view

PROTEIN STRUCTURE DETERMINATION

- Predict protein structures from their sequences with the novel AI/ML algorithms OpenFold[1] and AlphaFold[2]
- Predict structures of protein sequences co-folded with ligands, nucleic acids, and peptides with OpenFold
  - Co-fold inputs simultaneously, or screen libraries of inputs one-by-one
- Generate high-quality 3D models of target proteins from their sequences with the market leading MODELER[3] homology modeling algorithm
- Assess the model quality with tools including, model confidence, scoring functions, energies, and sequence-structure compatibility
- Prepare protein structures for molecular dynamics and docking studies using a comprehensive set of automatic protein preparation tools
  - Removes disorder and extraneous water molecules
  - Detects and adds missing atoms and residues with additional refinement
  - Predicts pKas of titratable amino acids and protonates at the desired pH for optimal interactions[4]

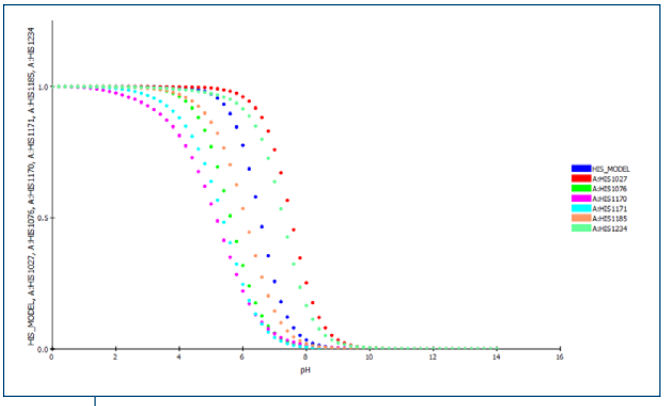


Figure 3: Analyze titration curves of prepared residues

REFINEMENT AND SIMULATION

- Systematically sample and refine loop conformations using the CHARMm-based LOOPER[5] algorithm
- Graft loop conformations from a template structure onto a target model
- Optimize amino acid side-chain positions using the CHARMm-based ChiRotor[6] algorithm
- Prepare proteins in an explicit membrane with solvation for Molecular Dynamics (MD) simulations
- Perform GPU-enabled explicit solvent MD simulations with CHARMm[7], OpenMM[8] or NAMD[9] to explore conformational space and measure structural and energetic properties
- Apply the Gaussian accelerated Molecular Dynamics (GaMD)[10] enhanced sampling method to accelerate sampling of protein conformations
- Analyze MD trajectories to investigate binding pocket dynamics and structural features of allosteric motions
- Examine electronic effects in protein-ligand complexes using a hybrid of quantum and classical molecular mechanics (QM-MM)

PROTEIN-PROTEIN DOCKING

- Predict the structures of protein-protein complexes quickly and accurately with ZDOCK[11]
  - Cluster poses based on their spatial proximity and filter poses based on known interface residues
- Refine docked poses with RDOCK to optimize binding interactions
- Analyze protein binding interfaces and generate reports for different types of interactions

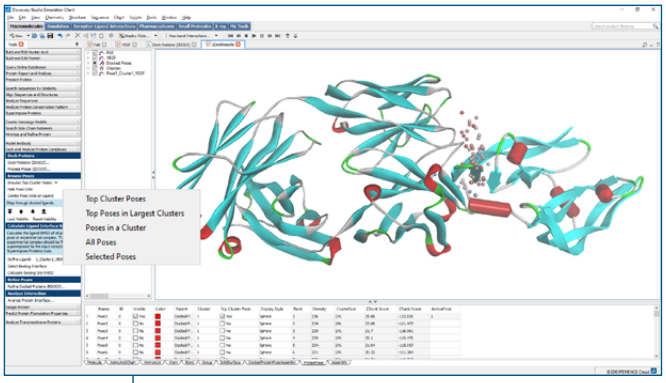


Figure 4: Analyze docked protein poses

PROTEIN DESIGN AND ENGINEERING

- Perform *de novo* design of protein binders for a protein target
  - Automatically generate new protein scaffolds, assign sequences and build structures combining the AI tools RFdiffusion[12], ProteinMPNN[13] and AlphaFold

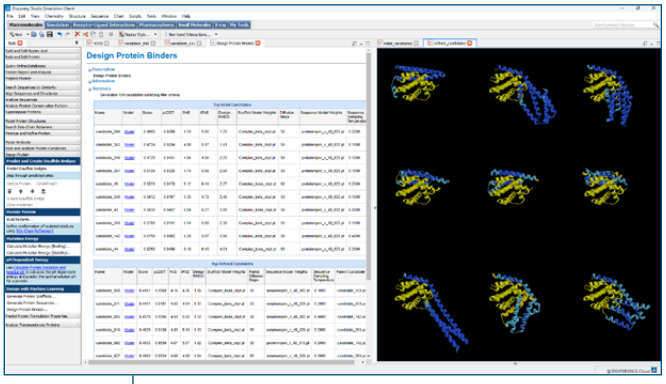


Figure 5: Design protein binders using a sequential optimization process

- Perform combinatorial amino acid mutagenesis to evaluate the effects of mutations on protein stability and binding affinity, considering pH dependency and thermal effects[14,15]
  - Mutations can be to a single residue such as in alanine scanning or in selected, complex combinations
  - Mutations can be multi-site mutations to identify the optimal mutation combination for protein binding or stability
  - Clear analysis about the predicted effect of the mutation and how it is composed is provided
- Identify mutation sites for disulfide bridge creation to improve protein stability

Index	Mutation	Mutation Energy (kcal/mol)	Effect
1	I:GLY29>LYS	-2.72	STABILIZING
2	I:GLY29>HIS	-1.29	STABILIZING
3	I:GLY29>ARG	-1.05	STABILIZING
4	I:CYS3>LYS	-0.93	STABILIZING
5	I:ARG1>HIS	-0.64	STABILIZING
23	I:PRO4>LYS	2.20	DESTABILIZING
24	I:ARG5>LYS	2.61	DESTABILIZING
25	I:ILE6>LYS	2.94	DESTABILIZING
26	I:PRO4>ARG	3.39	DESTABILIZING
27	I:ARG5>HIS	3.78	DESTABILIZING

The table reports up to 5 lowest energy and up to 5 highest energy mutations. For the full list of results click the links in the Results section.

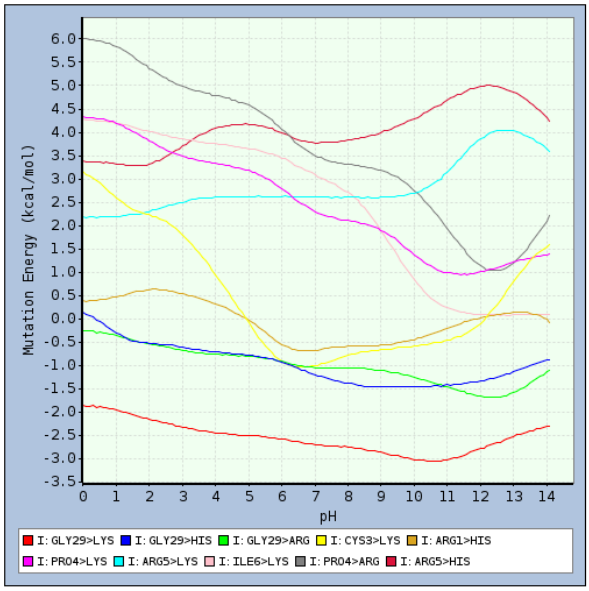


Figure 5: A summary of the lowest and highest energy mutations and the corresponding effect of the mutation. A line plot of mutation energy against pH is also available for these mutations.

- Identify mutation sites for disulfide bridge creation to improve protein stability

PROPERTY PREDICTION

- Calculate biophysical properties, such as solubility, isoelectric point, dipole moment, molecular charge, molar extinction coefficient, hydropathy and antigenic sites
- Calculate protein features and sequence descriptors for use in machine learning applications

FUNDAMENTAL MODELING TOOLS

- Rapidly build peptide molecules in defined secondary structure conformations
- Easily create RNA and DNA molecules in single or multi-stranded conformations according to standard A, B and Z forms
- Quickly assess structures from the RCSB with detailed reporting and analysis tools
- Specify the preparation of a protein, including standardize atom names, select alternate conformations, insert missing main-chain or side-chain atoms, adjust terminal residues, and more
- Examine backbone conformations of residues graphically with interactive Ramachandran plots for structure validation
- Align and superimpose protein structures based on structural or sequence similarity
  - Detailed RMSD analysis available at the residue level
- Perform simple x-ray structure determination and model structure refinement with CNX (Crystallography and NMR Explorer)

LEARN MORE

REFERENCES

1. https://openfold.io/
2. Jumper, J.; Evans, R.; Pritzel, A. *et al. Highly accurate protein structure prediction with AlphaFold.* Nature 2021, 596, 583–589.
3. Sali, A.; Potterton, L.; Yuan, F. *et al. Evaluation of comparative protein modeling by MODELLER.* Proteins 1995, 23, 318-326.
4. Spassov, V.Z.; Yan, L.; *A fast and accurate computational approach to protein ionization*. Prot. Sci. 2008, 17, 1955–1970.
5. Spassov, V.Z.; Flook, P.K.; Yan, L. *LOOPER: A molecular mechanics-based algorithm for protein loop prediction.* Prot. Eng., Design & Selection 2008, 21, 91-100.
6. Spassov, V.Z.; Yan, L.; Flook, P.K. *The dominant role of side-chain backbone interactions in structural realization of amino acid code. ChiRotor: A side-chain prediction algorithm based on side-chain backbone interactions.* Prot. Sci. 2007, 16, 494-506.
7. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D. *et al. CHARMM: A program for macromolecular energy minimization and dynamics calculations.* J. Comp. Chem. 1983, 4, 187-217.
8. https://openmm.org/
9. Phillips, J.C.; Braun, R.; Wang, W. *et al. Scalable molecular dynamics with NAMD.* J. Mol. Biol. 2005, 26, 1781-1802.
10. Miao, Y.; Feher, V.A.; McCammon, J.A. *Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation.* J. Chem. Theory Comput. 2015, 11, 3584–3595.
11. Chen, R.; Li, L.; Weng, Z. *ZDOCK: An initial-stage protein-docking algorithm.* Proteins 2003, 52, 80-87.
12. Watson, J.L.; Juergens, D.; Bennett, N.R. et al., *De novo design of protein structure and function with RFdiffusion.* Nature 2023, 620, 1089-1100.
13. Dauparas, J.; Anishchenko, I.; Bennett, N. et al. *Robust deep learning–based protein sequence design using ProteinMPNN*. Science 2022, 378, 49-56.
14. Spassov, V.Z.; Yan, L.; *pH-selective mutagenesis of protein-protein interfaces: in silico design of therapeutic antibodies with prolonged half-life.* Proteins 2013, 81, 704-14.
15. Spassov, V.Z.; Yan, L.; *A pH-dependent computational approach to the effect of mutations on protein stability.* J. Comput. Chem. 2016, 37, 2573-87.

Dassault Systèmes is a catalyst for human progress. Since 1981, the company has pioneered virtual worlds to improve real life for consumers, patients and citizens.

With Dassault Systèmes' **3D**EXPERIENCE platform, 370,000 customers of all sizes, in all industries, can collaborate, imagine and create sustainable innovations that drive meaningful impact.

For more information, visit: **www.3ds.com**

Virtual Worlds for Real Life

DASSAULT SYSTEMES

DS-10043-0326